

## From Text to Terminology ... and Back to Text



Sylvia Ball

SILD Division, European Parliament

### 1. Introduction

As a subtitle to my presentation for the LSP conference I added the phrase “terminology management as an aid to translation quality and timeliness in an expanding European Union”, because what I set out to do was to situate terminology activity at the European Parliament (EP) within the institution's role and the translation processes necessary for it to work effectively in eleven languages at present and 21 or more in the future. This paper will accordingly describe the role of the EP today and the translation and terminology tools used within its Translation Service, look ahead to the future, insofar as it is possible to do so and suggest a few tentative conclusions.

The EP has been active in the field of terminology for over 35 years. Our Terminology Office (BdT) of the 1960s and 1970s published glossaries, journals and other works, which were well thought of but not always directly relevant to the type of text drafted and translated at the EP<sup>1</sup>. As parliamentary activity has expanded over the years with a corresponding increase in the number of languages and the translation workload, it has become far more important to manage relevant terminology proactively, to ensure that correct terminology is available to European Union translators (and others involved in the text drafting and translation process) in a timely fashion. This is the main current aim of the terminology team within the EP's SILD Division<sup>2</sup>, working on both traditional and automatic term extraction.

---

<sup>1</sup> For more detail about this topic, see Ball (1997).

<sup>2</sup> SILD = *Support informatique, linguistique et documentaire* (IT, Language and Documentation Support)

## 2. The European Parliament

The European Parliament is the European Union's democratic institution, elected by universal adult suffrage every five years, although there is not yet a uniform electoral procedure for the fifteen countries of the Union or even a uniform election date. The next elections will take place in June 2004, by which time it is expected that the fifteen will have increased to 25 and the current eleven languages to twenty<sup>3</sup>. The five-year parliamentary term is broken down into two halves, with the Bureau, our principal executive structure, and other officers such as committee chairmen re-elected at mid-term, a fact which can have an impact on its Translation Service at the most basic level, as it did in January 2002 when an Irish male President, Pat Cox, succeeded Mrs Nicole Fontaine, who had been the first female EP President to opt to be “Madame la Présidente”. All our macros had to be adapted accordingly, in all the languages which distinguish two forms<sup>4</sup> although we have kept the feminine variants for future reference.

For the record, the EP's annual session theoretically runs from the second week of March<sup>5</sup>, although our budgetary experts (the EU budget being historically the institution's first area of real authority) would have us believe that the parliamentary year, like an academic year, begins in the autumn, which is dominated by budgetary procedure to the point of a whole part-session being devoted to it. Twelve part-sessions per year are held in Strasbourg, each lasting four days, with additional part-sessions (universally known as “mini-part-sessions” because they last for the afternoon of one day and the morning of the next) held in Brussels. Part-session plenary procedure is linked to committee meetings held in Brussels (or in Strasbourg during part-session weeks). Brussels is also the venue for most meetings of the multinational (and thus almost inevitably multilingual) political groups. The institution's secretariat is split between Brussels and Luxembourg with most staff who work in direct contact with MEPs being based in the former and some 2000 others, including the Translation Service proper<sup>6</sup>, based in the latter.

---

<sup>3</sup> For more detail about the EP, its electoral procedure and parliamentary procedure in general, see *Fact Sheets*, URL: [http://www.europarl.eu.int/factsheets/default\\_en.htm](http://www.europarl.eu.int/factsheets/default_en.htm)

<sup>4</sup> And, to a lesser extent in English, also, to take account of standard phraseology such as “instructs him/her to forward ...”

<sup>5</sup> See European Parliament, *Rules of Procedure*, Rule 10(2).

<sup>6</sup> For the purposes of this paper the Translation Service excludes the Freelance Unit, although it is in the same Directorate-General, the Parliamentary Acts Division, responsible for the Part-Session Minutes and the Verbatim Report of Proceedings, the Lawyer-Linguists Unit, based in Brussels and the proposed Tabling Office, which will provide support to MEPs throughout an individual parliamentary procedure.

Parliamentary procedures are essentially similar to those in many other democratic assemblies, involving a variable number of readings, amendment stages, etc., with the important distinction that in the EP they are all multilingual. Our Rules of Procedure specifically provide that, for a text to be valid, it must be available in all language versions<sup>7</sup> and, while it is not a very frequent occurrence, it is not unknown for the proceedings of the plenary session to be held up while a ruling is given on linguistic interpretation. Interestingly, it is not always the language in which a text was originally drafted which is the version finally adopted. Hence the importance of having a service staffed by professional translators all working into their mother tongue.

### **3. The Translation Service**

The Translation Service is currently divided into eleven language divisions each with between 30 and 45 translators plus approximately 20 support staff (secretaries, archivists, etc). Across-the-board support in the areas of translation tools and IT tools in general, terminology management and documentation is provided by SILD, working in coordination with translation division staff with expertise in particular areas and with the EP's central IT Directorate. Compared with its potential client base the SILD Division is very small, with twelve linguists (including the Head of Division whose role is essentially managerial), some additionally with specific IT qualifications, an IT development team of 5-6, a hands-on support unit of a similar size and a small number of specialist support staff (8-10).

As might be expected in an institution where the number of MEPs per country varies from six (for Luxembourg) to 99 (for Germany) and where a small number of the languages concerned are disproportionately widely used for international communication, the source-language (SL) distribution of texts for translation is extremely variable, although all translation divisions have translators able to work from all the other languages. (Most experienced translators work from four or five SLs.) In the year 2002 the SL distribution was as follows: English 40%, French 25%, German slightly under 10%, Spanish 5.95% and Italian 4.6%, leaving only a little over 14% for the other six languages! Twenty years ago, of course, the vast majority of session documents were drafted in French (although not necessarily by native French speakers), with English, German and Italian very much “also-rans”. However, the distribution of target-language pages is much more homogeneous with (again

---

<sup>7</sup> See *Rules of Procedure*, Rule 117.

for 2002) between 53000 and 69000 pages<sup>8</sup> per language. Unsurprisingly, English is now the language with the lowest figure, but the highest is for German, closely followed by French, which goes to show how many non-legislative texts are translated into relatively few languages.

One of the main impacts of parliamentary business on the Translation Service comes from the fact that multiple versions of legislative texts are the norm rather than the exception, as documents progress from one stage of procedure to the next, with frequent changes of source language along the way. This can obviously make continuity more difficult, since the translator responsible for the earlier versions of a text may not necessarily translate from the language of the final version. However, it goes a long way to making my point that we extract terminology from texts and then recycle it back into future versions.

This is also due, of course, to the fact that the EP does not work in isolation, but in conjunction with the other European institutions, giving its opinion on proposals from the European Commission or acting as a co-legislator with the Council of the EU. With the exception only of own-initiative reports and areas where the Treaties give the EP the right of initiative, there is always a prior text (with terminology to be recycled - or occasionally rejected) on which the EP text is based. Such texts are themselves based on our founding Treaties and other binding documents, giving our terminology its authenticity.

#### **4. Translation Tools**

To make all staff's tasks easier, the central IT directorate supplies us with PCs equipped with the software tools that we have come to expect as part of any office worker's life: word-processors (in our case Microsoft Word 97), spreadsheets (Excel), e-mail, internet access, etc. For Translation Directorate staff, both translators and secretaries, SILD further provides the standard software that is needed for their specific tasks: a terminology database management system and associated translation tools, in our case the Trados 5.5 suite of applications, with an add-on called "CATMan", designed to integrate the individual Trados tools so that translators can concentrate on their core task. To improve standardisation of texts (both originals and translations), we also play a part in developing and maintaining Word macros and XML-based tools for standard texts in general<sup>9</sup> and legislation in particular<sup>10</sup>. Legisdoc, the latter application, is designed to be used with the Translator's Workbench (TWB) and

---

<sup>8</sup> A page is defined as 300 words or 1500 characters (not including spaces).

<sup>9</sup> DocEP is a suite of Word macros for the production of standard texts.

<sup>10</sup> Legisdoc is a system for the production of multilingual legislative texts in XML.

Word or its successor applications to ensure that all language versions of texts are consistent and to speed up publication whether on Europarl or in hard-copy. However, since many experienced translators prefer to dictate their texts (having had no option but to do so for much of their career), we also provide support for two speech-recognition packages, Dragon NaturallySpeaking, which is used by several translation divisions, and Philips Viva Voice, used in the Portuguese division. For translation-specific applications SILD support is not limited to equipment supply, organising training and trouble-shooting, since nowadays clients expect more. The support staff for the Translator's Workbench (TWB) and Dragon speech-recognition software manage intranet sites for their specific sectors with the usual features: basic instructions, tips and tricks, links to relevant internet sites, FAQs, etc.

## **5. Terminology Support and Tools**

The terminology team within SILD is quite small, comprising two linguists (one English and one French mother-tongue), one secretary and three to four trainees at any one time. Increasingly, trainees come from the countries which are expected to join the EU in 2004-2007, although this is not the only factor in their selection<sup>11</sup>. In addition to mentoring trainees and recording and collating terminology in their working languages (in Trados MultiTerm 5.5 for maximum compatibility with the tools used by translators<sup>12</sup>), the SILD terminologists act as coordinators for terminology activity in the translation divisions.

Cooperation may be quite structured, as in the case of the ongoing project to analyse and record the terminology used in the Official Journal of the European Union, or ad hoc and informal, as in the case of the terminology recorded by translation division staff in the interdivisional terminology database which feeds into the institution's main terminology database EUTERPE<sup>13</sup>, or in the reports compiled by translators who observe committee meetings in Brussels and analysed by the terminology team. In the latter case it could be said that we are even going beyond “from text to terminology”, since the reports sometimes give an insight into the sub-text, the reasoning behind certain terminology preferences or dislikes. This adds an extra dimension to certain terminology records.

---

<sup>11</sup> For more information about traineeships in the Translation Service, see Europarl, URL: [http://www.europarl.eu.int/stages/default\\_en.htm](http://www.europarl.eu.int/stages/default_en.htm)

<sup>12</sup> For EP staff outside Translation or those unable/unwilling to work in MultiTerm, we also provide an intranet terminology proposal form for further processing by terminologists.

<sup>13</sup> EUTERPE = *Exploitation unifiée de la terminologie au Parlement européen* or European Parliament one-stop terminology management system

EUTERPE, to date<sup>14</sup>, has a little over 275000 records in one to eleven EU languages, often with abbreviations or acronyms, sometimes with Latin (for scientific terms) and, occasionally, with terms from non-EU languages, including approximately 2000 terms in Norwegian. The standard wording in the opening record states that:

The EUTERPE database contains 275,007 entries from the following areas: EU terminology as found in the Official Journal, including projects and programmes; Rules of Procedure of the EP and other institutions; EP bodies; Member States' national, local and regional government, and political parties; acronyms and abbreviations; currencies and geographical data; EDP, telecommunications and online communication; science, technology and statistics; AIDS, medicine and health; botany and zoology; social policy including asylum; education, training and qualifications; environmental issues, ecology and pollutants; transport and air-traffic control.

Many of the corpora included in EUTERPE are the results of projects with our trainees or the translation divisions, but in recent years there has been less emphasis on thematic projects and more on the analysis of the terminology used in the Official Journal of the EU, because that gives the best results for translation divisions. For translators who work directly in Microsoft Word, taking terms from EUTERPE is literally a case of “back to text”, since double-clicking on the relevant term pastes it into the translation at the point required. Although the SILD division has recently acquired licences for the Xerox Terminology Suite, at the present time terminology extraction is almost always manual, based on the solutions found by translators to the problems they have encountered in their work or on the SILD's analysis of the OJ, official documents or the press. Progress with the introduction of automated terminology extraction has so far been slow, largely because the results are inferior to traditional methods, although one of our current trainees has undertaken to test the Xerox software in Spanish and English and to compile a report, so that we can see how to get better results. Nonetheless, given the speed with which technology advances, it seems safe to assume that within the next three to five years automatic terminology extraction software will become a standard tool.

The terminology team and other members of the SILD division also manage a number of terminology-related intranet pages with links to other databases<sup>15</sup> or

---

<sup>14</sup> Mid-March 2003

<sup>15</sup> TIS, the Council of the EU's Terminological Information System, Eurodicautom, the Translation Centre's Euroterms, Eurocontrol's Euronav, etc.

sources of information including CELEX, the interinstitutional full-text database of Community law. However, each database has its own structure and search procedures (and, since much of the terminology is binding on us all, there is a vast amount of overlap between them) so, to rationalise and harmonise terminology activity between the EU institutions, in 1999 it was decided to merge the existing databases and provide a single terminology database for the European Union in what has become known as the IATE project<sup>16</sup>.

## **6. The IATE Project**

The project originated in the Translation Centre for the Bodies of the European Union, set up in 1993 to translate the texts produced by the various decentralised bodies set up at that time or which already existed. Hence the acronym's first full form, the "Inter-Agency Terminology Exchange". Once it became an interinstitutional project a new name, reflecting its wider ambitions was needed and, since one of the main aims is to get away from the rather static databases of the 1990s, where only a few insiders saw terminology as soon as it was recorded, "Inter-Agency" gave way to "Interactive". Again, since all the institutions taking part regard it as vital that it be a single platform for terminology activity in their translation services, it clearly had to be more than just an "exchange". At a time when the EU is gearing up for its greatest expansion yet, why not go continent-wide by adding "for Europe"?

More seriously, the new system is designed to push terminology activity "up-stream" to translation divisions in the institutions where terms originate, wherever possible, thus making sure that duplication of effort is minimised and maximum relevancy is achieved. This will give translators at the EP, working on texts relating to those from the Commission or Council, timely access to the terminology previously used in the other institutions, with the most economical use of resources in the central terminology unit. In the future, in addition to extracting terminology using automated tools, terminologists will concentrate on validation and terminology administration and diffusion. They will be able to communicate with the translation-division staff who work with them by using tools built into the IATE system<sup>17</sup>. There are also specific tools and routines for managing formal and content-validation cycles.

---

<sup>16</sup> IATE = Interactive Terminology for Europe. For the history of the IATE project, see MacPhail (2000), Rummel/Ball (2001) and Almeida (2001).

<sup>17</sup> See Rummel/Ball (2001) for more information about the communication modules, called "Messaging" and "Marks".

With an eye to 2004 and beyond, the system is designed to be scalable. As translators arrive for the new languages they will be able to add the equivalents decided on for the core EU terminology from the “acquis communautaire”, on the fly or by batch uploads of new terms<sup>18</sup>. Since everybody will be working on a single platform there should be no duplication of effort and, since all users will be able to communicate with each other, there should be maximum harmonisation. As a further benefit, the IATE system was designed from the outset to be Unicode compliant so that it should make no difference if users need to enter terms or ancillary data with the many and varied special characters of the new languages, or in Cyrillic (for Bulgarian or for Russian, used for some Translation Centre texts).

This all sounds quite perfect, but IATE has not always been a “long fleuve tranquille”. Apart from purely practical problems such as the fact that the system should have gone live in July 2001 and is now not expected to enter the production phase until July 2003, all institutions have had to accept their interdependence and learn to work together and, most importantly, to overcome their fear of the surface complexity of the new system. For MultiTerm users at least, accustomed to simple interfaces and relatively few options, initial reactions ranged from reserved acceptance to total refusal. Indeed, one explanation of the project delays has been the need to make the system more user-friendly, by offering more basic interfaces for translators than for experienced terminologists, willing and able to take the time to research the variety of information that the system allows for. A translator who has solved a terminology problem cannot necessarily afford the time to draft a definition (perhaps in both source and target languages) or scope note. All we can expect is that s/he will enter the source term, the equivalent and the reference to the source for both, perhaps with the source context if it helps illustrate the concept.

The changeover has also forced all the participants to look at their own practices critically. Certainly, at the EP, it has shown many areas where we were inconsistent or over-subjective (in particular in marking abbreviations as “belonging” to the wrong language, but also in the use of fields such as “context”, which sometimes contained irrelevant information), and several other participants in the project have reported similar findings. It is to be hoped that the data finally loaded into the IATE system will benefit from the experience, and that the result will be worth the long delay.

---

<sup>18</sup> The European Commission is working on an automated system developed in-house for this feature.



## 7. The Future

As I pointed out at the beginning of this paper, it is always hazardous to try to second-guess the future. However, particularly now that the referendum process has got off to such a successful start with the Maltese vote in favour of accession, what we are looking towards is an EU of 25 countries with 20 working languages and 732 MEPs from 1 May 2004. To serve the democratic interests of the citizens of the enlarged EU, the Translation Service will expand to 20 translation divisions, with 3 pivot divisions (English, French and German) each ideally covering all the other 19 languages and another 17 divisions covering at least the pivot languages in the first instance and, in the longer term, Spanish and Italian as well<sup>19</sup>.

To cope with the increased workload, the Translation Service will adopt a system of controlled multilingualism, where all languages can be used for text drafting and all legislative texts will be available in all languages, but some translations will be the result of a two-stage operation, first into a pivot language and then into the target language. In the longer term, again, the new languages which are expected to be those which will generate the most source language texts, viz. Polish, Hungarian and Czech, will be covered by all the pre-2004 translation divisions. There will also be a gradual increase in the use of external translation, although not all EP texts, legislative ones in particular, are amenable to contracting-out.

As the profile of staff changes with the recruitment of translators who take computer literacy for granted and expect to work on-screen, the tasks of secretaries and other support staff will also change. They will evolve into “translation assistants”, working on pre-translations using aligned texts, checking alignments in their working languages<sup>21</sup> and, no doubt, extracting terminology and creating or updating IATE records for checking by a translator or terminologist.

It sounds optimistic and, no doubt, in reality it will not all be quite so easy. It never is. But the previous enlargements have shown us that, with good will on all sides, enlargements do work. The years to come will certainly be, in Robert Kennedy's famous phrase, “interesting times”!

---

<sup>19</sup> The staffing levels envisaged for the pivot-language divisions are 40 translators + 20 support staff, with 35 translators + 18 support staff for the others.

## References

- Almeida, António (2001) Die Terminologische Datenbank der Europäischen Union. In Braun, Angelika (Hrsg.) *Festschrift Joachim Göschel, Beiträge zu Linguistik und Phonetik*. Supplement to the *Zeitschrift für Dialektologie und Linguistik*. Stuttgart.
- Ball, Sylvia (1997) In the Beginning was the Glossary: the development of integrated language support services at the European Parliament. *Terminologie et Traduction* (2/1997).
- European Parliament (2003) *Fact Sheets*. URL [http://europarl.eu.int/factsheets/default\\_en.htm](http://europarl.eu.int/factsheets/default_en.htm)
- European Parliament (2003) *Rules of Procedure*. Fifteenth Edition. Luxembourg: Office for Official Publications of the European Communities.
- MaPhail, Alistair (2000) IATE - Inter-Agency Terminology Exchange. *Conference for a Terminology Infrastructure in Europe*. Paris.
- Rummel, Dieter/Ball, Sylvia (2001) The IATE Project - Towards a Single Terminology Database for the European Union. *Translating and the Computer*. London: Aslib